DOCUMENT RESUME

ED 466 497 TM 034 236

AUTHOR La Marca, Paul M.; Redfield, Doris; Winter, Phoebe C.

TITLE State Standards and State Assessment Systems: A Guide to Alignment. Series on Standards and Assessments.
INSTITUTION Council of Chief State School Officers, Washington, DC.
SPONS AGENCY Office of Elementary and Secondary Education (ED), Washington, DC.
PUB DATE 2000-00-00
NOTE 36p.; Sponsored by State Collaborative or Assessment and Student Standards, Comprehensive Assessment Systems for IASA Title I. Written with assistance from Adrienne Bailey and Linda Hansche Despriet.
AVAILABLE FROM Council of Chief State School Officers, Attn: Publications, One Massachusetts Ave., N.W., Suite 700, Washington, DC 20001 ($10). Tel: 202-336-7016; Fax: 202-408-8072; Web site: http://www.ccsso.org/publications.
PUB TYPE Guides - Non-Classroom (055) -- Reports - Descriptive (141)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Standards; Curriculum; Educational Assessment; Elementary Secondary Education; Performance Factors; Policy Formation; State Programs; *State Standards; *Testing Programs
IDENTIFIERS *Curriculum Alignment

ABSTRACT

Alignment of content standards, performance standards, and assessments is crucial. This guide contains information to assist states and districts in aligning their assessment systems to their content and performance standards. It includes a review of current literature, both published and fugitive. The research is woven together with a few basic assumptions, best practice, and practical reality to produce a resource for planning and achieving a comprehensive aligned system of standards and assessments. The guide rests on six general assumptions about the foundations of an aligned system and draws on relevant research in discussing critical aspects of alignment: (1) content match; (2) depth match; (3) emphasis; (4) performance match; (5) accessibility; and (6) reporting. It also discusses alignment in the context of other components of the educational system, including accountability, teacher involvement and professional development, policy development, textbook adoption and use, and K-16 connections. An appendix contains an annotated checklist that states and localities can use to evaluate the degree of alignment of their assessments and standards or to develop an aligned assessment system. (Contains 28 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made from the original document.

TM

CAS
COMPREHENSIVE ASSESSMENT SYSTEMS FOR TITLE I

ED 466 497

State Standards and State Assessment Systems: A Guide to Alignment

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
B. Buterbaugh
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

Paul M. La Marca
Nevada Department of Education

Doris Redfield
AEL, Inc.

Phoebe C. Winter
Council of Chief State School Officers

with

Adrienne Bailey
Council of the Great City Schools

Linda Hansche Despriet
Georgia State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM034236

Council of Chief State School Officers
State Collaborative on Assessment and Student Standards

BEST COPY AVAILABLE

State Standards and State Assessment Systems: A Guide to Alignment

Paul M. La Marca
Nevada Department of Education

Doris Redfield
AEL, Inc.

Phoebe C. Winter
Council of Chief State School Officers

with

Adrienne Bailey
Council of the Great City Schools

Linda Hansche Despriet
Georgia State University

Council of Chief State School Officers
One Massachusetts Avenue, N.W.
Washington, DC 20001

© Copyright 2000 by the Council of Chief State School Officers. This document may not be reproduced without the permission of the Council of Chief State School Officers, the copyright holder.

The Council of Chief State School Officers (CCSSO) is a nationwide, nonprofit organization composed of the public officials who head departments of elementary and secondary education in states, the District of Columbia, the Department of Defense Education Activity, and five extra-state jurisdictions. CCSSO seeks its members' consensus on major education issues and expresses their views to civic and professional organizations, to federal agencies, to Congress, and to the public. Through its structure of standing committees and special task forces, the Council responds to a broad range of concerns about education and provides leadership on major education issues.

Because the Council represents each state's chief education administrator, it has access to the educational and governmental establishment in each state and to the national influence that accompanies this unique position. CCSSO forms coalitions with many other education organizations and is able to provide leadership for a variety of policy concerns that affect elementary and secondary education. Thus, CCSSO members are able to act cooperatively on matters vital to the education of America's young people.

The State Education Assessment Center is a permanent, central part of the Council. The Center was established through a resolution by the membership of CCSSO in 1984. This report is part of a series sponsored by the Assessment Center's State Collaborative on Assessment and Student Standards (SCASS), Comprehensive Assessment Systems for IASA Title I. The series addresses issues related to the standards and assessment provisions of Title I.

Preparation of this report was supported in part by the U.S. Department of Education, Office of Elementary and Secondary Education. The views and opinions expressed in this report are not necessarily those of the Council of Chief State School Officers, the U.S. Department of Education, or the State Collaborative on Assessment and Student Standards.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

GORDON M. AMBACH
EXECUTIVE DIRECTOR

WAYNE N. MARTIN
DIRECTOR
STATE EDUCATION ASSESSMENT CENTER

JOHN F. OLSON
DIRECTOR OF ASSESSMENTS
STATE EDUCATION ASSESSMENT CENTER

PHOEBE C. WINTER
PROJECT DIRECTOR, COMPREHENSIVE ASSESSMENT SYSTEMS FOR IASA TITLE I
STATE COLLABORATIVE ON ASSESSMENT AND STUDENT STANDARDS

State Collaborative on Assessment and Student Standards
Comprehensive Assessment Systems for IASA Title I

SERIES ON STANDARDS AND ASSESSMENTS

Opportunity and Change: An Overview of Title I of the Improving America's Schools Act of 1994. Frank Philip, 1995

Designing Coordinated Assessment Systems for Title I of the Improving America's Schools Act. Edward Roeber, 1995

Adequate Yearly Progress Provisions of Title I of the Improving America's Schools Act: Issues and Strategies. Dale Carlson, 1996

Implementing the Adequate Yearly Progress Provisions of Title I in the Improving America's Schools Act of 1994. Phoebe Winter, 1996

Using Existing Assessments for Measuring Student Achievement: Guidelines and State Resources. Linda Hansche, Tom Stubits, and Phoebe Winter, 1997

Analyzing, Disaggregating, Reporting, and Interpreting Students' Achievement Test Results: A Guide to Practice for Title I and Beyond. Richard Jaeger and Charlene Tucker, 1998

Handbook for the Development of Performance Standards: Meeting the Requirements of Title I. Linda Hansche, with contributions from Ronald Hambleton, Craig Mills, Richard Jaeger, and Doris Redfield, 1999

Primary Level Assessment for IASA Title I: A Call for Discussion. Diane M. Horm-Wingerd, Phoebe C. Winter, and Paula Plofchan, 2000

PREFACE

This report is the work of the State Collaborative on Assessment and Student Standards, Comprehensive Assessment Systems for IASA Title I (SCASS CAS) study group on aligning assessments to standards. It is intended to provide guidance to states and localities as they develop or refine their assessment systems or evaluate the alignment of their assessment systems to their content and performance standards. We hope that this report also serves as a resource for informing policymakers of the critical importance of an aligned assessment system to the overall educational system.

SCASS CAS ALIGNMENT STUDY GROUP

Paul LaMarca (Chair)	Nevada Department of Education
Adrienne Bailey	Council of Great City Schools
Rolf Blank	Council of Chief State School Officers
James Grissom	California Department of Education
Doris Redfield	AEL, Inc.
Collette Roney	U.S. Department of Education
Grace Ross	U.S. Department of Education
Elois Scott	U.S. Department of Education
Tom Stubits	Pennsylvania Department of Education
David Sweet	U.S. Department of Education
Cheryl Tibbals	Council of Chief State School Officers
Kit Viator	Massachusetts Department of Education
Hugh Walkup	U.S. Department of Education
Phoebe Winter	Council of Chief State School Officers

The authors gratefully acknowledge Bill Erpenbach's and John Olson's reviews and comments on an earlier draft of this document.

EXECUTIVE SUMMARY

Alignment of content standards, performance standards, and assessments is crucial, whether assessment data are used to make policy decisions, funding allocations, or recommendations for student learning. This guide contains information to assist states and districts in aligning their assessment systems to their content and performance standards. It includes a review of current literature, both published and fugitive. This research is woven together with a few basic assumptions, best practice, and practical reality to produce a resource for planning and achieving a comprehensive aligned system of standards and assessments.

The guide rests on six general assumptions about the foundations of an aligned system of standards and assessments:

1. An aligned system of standards and assessments will meet its goal of improving student performance only if curriculum is also part of the aligned system.
2. In an aligned system of standards and assessments, classroom instructional practices must be based on and clearly reflect the content standards and curriculum.
3. Alignment of state assessments to state standards will depend on the alignment of educational practices and philosophies between state and local education agencies.
4. State standards and state assessments should be visible and unguarded.
5. Alignment should be viewed as an ongoing process in need of periodic evaluation and adjustment.
6. Valid and meaningful data-based decision-making depends on the degree of alignment between standards and assessments.

The guide draws on relevant research findings in discussing critical aspects of alignment: content match, depth match, emphasis, performance match, accessibility, and reporting. It also discusses alignment in the context of other components of the educational system, including accountability, teacher involvement and professional development, policy development, textbook adoption and use, and K–16 connections. An Appendix to the guide contains an annotated checklist that states and localities can use as a resource for evaluating the degree of alignment of their assessments and standards or for developing an aligned assessment system.

TABLE OF CONTENTS

PREFACE i

EXECUTIVE SUMMARY iii

I. INTRODUCTION 1

DEFINITIONS 1
WORKING ASSUMPTIONS CONCERNING ALIGNMENT 2
Standards Development 4
Assessment Development 5
Reporting 6
OTHER ISSUES RELATED TO ALIGNMENT 7
Federal Program Requirements 7
Standards of National and Professional Associations 8
Norm-Referenced and Criterion-Referenced Tests 8
Multiple Measures 9
SUMMARY 9

II. DEVELOPING AND EVALUATING SYSTEM ALIGNMENT 11

GENERAL ORGANIZING PRINCIPLES 11
Content Match 11
Depth Match 11
Emphasis 12
Performance Match 12
Accessibility 13
METHODOLOGY 15
SUMMARY 18

III. SUPPORTING MECHANISMS AND STRUCTURES OF AN ALIGNED SYSTEM FOCUSED ON STUDENT LEARNING 19

IV. CONCLUSION 21

APPENDIX: STANDARDS-ASSESSMENT ALIGNMENT CHECKLIST 23

PARTICIPANTS, TRAINING, AND MATERIALS 23
Participants 23
Training 23
Materials for Review 24
REVIEW CATEGORIES 24
Content Match 24
Depth Match 26
Emphasis 26
Performance Match 26
Accessibility 27
Reporting 27

REFERENCES 29

I. INTRODUCTION

Alignment of content standards, performance standards, and assessments is crucial whether assessment data are used to make policy decisions, funding allocations, or recommendations for student learning. Although alignment of these three major components makes intuitive sense, making it a reality is the challenge and the focus of this guide. This document is designed to provide state and local education agencies with a useful resource for addressing alignment issues.

The component parts of an aligned system are examined and methodology for achieving and evaluating alignment is explored in this guide, and it includes a review of current literature, both published and fugitive, as well as personal communication. This research is woven together with a few basic assumptions, best practice, and practical reality to produce a resource for planning and achieving a comprehensive aligned system of standards and assessments.

DEFINITIONS

Standards

Content standards specify the knowledge and skills students are expected to acquire through schooling in content areas such as reading and mathematics.

Performance standards specify the match between demonstrated knowledge of the content standards and specific proficiency levels. Hansche (1998) defines performance standards as consisting of four parts:

1. performance levels: labels for levels of achievement;
2. performance descriptors: descriptions of student performance at each level of achievement;
3. exemplars: illustrative student work for the range within each level; and
4. cut scores: score points on assessments that differentiate between performance levels.

Assessment

Assessment is a process that uses tests and/or other instruments and procedures to collect information that for use in making appropriate inferences about student learning. Included are tests or examinations used at all levels of the educational process, including assessments at the state, district, school, and classroom levels. This document focuses on large-scale assessments such as norm-referenced and standards-based assessments. Although levels of assessment other than large-scale are not dealt with in detail in this guide, the issues discussed are applicable to all levels of assessment.

Alignment

"Alignment is a match between two or more things. *Webster's New World College Dictionary* defines *align* as 'to bring into a straight line; to bring parts or components into a proper coordination; to bring into agreement, close cooperation.' In an aligned system of standards and assessments, all components are coordinated so that the system works toward a single goal: educating students to reach high academic standards" (Hansche, 1998, p. 21). Ultimately, alignment refers to how well all elements in a system work together to guide instruction and student learning (Webb, 1997a).

Alignment directly affects the degree to which valid and meaningful inferences about student learning can be made from assessment data (Long and Benson, 1998).

System Development

Simply stated, standards provide a framework for the knowledge and skills that students are expected to acquire (content standards) and how well they should be able to perform relative to the content standards (performance standards); assessments should provide information regarding the attainment of standards. Typically, system development follows a sequence in which content standards are developed first, followed closely by performance standards, and finally by the development of assessments, usually a lengthier process. Although developing assessments is not the best place to start, the beginning — standards development — cannot be completed until assessments are in place. Ideally, the components of the system can be developed in a joint or iterative fashion; content standards should be developed while considering performance standards and how those standards ultimately will be assessed. It is only in a fully aligned system that data with respect to student learning are meaningful.

Keep in mind that this document focuses on the alignment between standards and assessments. If content and performance standards specify the necessary knowledge, skills, and extent to which students can demonstrate knowledge and skills, assessments should provide an avenue to judge student achievement toward these ends.

Although standards and assessments form their own system, they are components of a larger "dynamic" educational system, meaning that the components are not static but evolve over time. Each component informs and is informed by other components; the degree of alignment between standards and assessments depends on system wide alignment. Other related matters to keep in mind include professional development, parent involvement, safety and discipline, technology, and school autonomy.

Working Assumptions Concerning Alignment

This guide is built on six working assumptions. Undoubtedly other key issues will affect the process of development and evaluation of the degree of system alignment as well. These six general assumptions, however, provide a sound foundation for consideration of system alignment.

1. *Though the emphasis of this document is on an aligned system of standards and assessments, the system will meet its goal of improving student performance only if curriculum is also part of the aligned system.*

   Most educators and parents are familiar with the concept of a curriculum. Content standards simply are descriptive statements about a given curriculum. Standards are not different from or in addition to curriculum. As Hansche states, "Think of a curriculum as a bridge, or conduit, between the broader vision of what is important in lay terms and what teachers should teach in their classrooms. The curriculum is simply an elaborated or 'technical' version of the content standards. Content standards and curricula are related tools; they do not contain different content to be learned, and [should] not [be] in conflict." (1998, p.22).

2. *In an aligned system of standards and assessments, classroom instructional practices must be based on and clearly reflect the content standards and curriculum.*

   The system cannot meet its goal of improved student performance if instruction is not aligned with the content standards and assessments. Standards and assessment may exhibit a high level of "face"

alignment, but the system becomes functional only when students are tested on information they have been taught, in a fashion similar to how the information was presented and learned. That information must be a direct reflection of the content standards and curriculum.

3. *Alignment of state assessments to state standards depend on aligning educational practices and philosophies between state and local education agencies.*

Across the country, states and local school districts are making substantial efforts to develop and implement challenging content standards and aligned assessments. The relationships between what states and local districts are doing have not received much attention. As states develop standards-based assessments, a consideration is how the state assessment fits with local assessment programs that may serve different purposes, use different instruments, and have different levels of commitment. It is becoming more and more important that the two systems support each other. While the decision rests at the state level on what components constitute an overall aligned system, the reality is that instruction occurs in the classroom and, therefore, local education agencies require some level of flexibility and must "buy in" to the state framework. The ability to base educational decisions on information garnered through state assessments will depend in large part on this macro-micro link.

4. *State standards and state assessments should be visible and unguarded.*

Information pertaining to state standards and state assessments should be accessible, understandable, and meaningful to all stakeholders, including policymakers, school administrators, the community at large, parents, and students (U.S. Department of Education, 1999). In particular, parents and students must be fully aware of state expectations regarding academic achievement and how state assessments will be used to judge or gauge achievement.

5. *Alignment should be viewed as an ongoing process in need of periodic evaluation and adjustment.*

All components of the system (standards, curriculum, instruction, and assessment) should be looked at as dynamic, not static, entities. The components are equally important and interdependent, and are engaged in a process that affects change in one another.[1]

6. *Valid and meaningful data-based decision making is dependent on the degree of alignment between standards and assessments.*

Ultimately, all of the issues, contexts, and implications addressed in this document are matters of validity. They all influence the extent to which results produced by a system of standards and assessments can be fairly and accurately used to address the purposes of the system. For example, if standards have been set and assessments developed as a system to determine which students will be promoted from one grade to the next,

1) the content standards must clearly and accurately represent the knowledge and skills, including cognitive skills, that students are expected to achieve as a function of schooling;
2) the performance standards must clearly, fairly, and accurately reflect multiple predetermined levels of performance relative to the content standards; and
3) in a perfectly aligned system, the assessments would cover the breadth and depth of both content and performance standards.

1 There certainly are other components of the educational system that require attention, and that affect and are affected by issues of alignment (e.g. professional development activities, textbook adoption, parental involvement, and the K–16 connection.) These and other related components are discussed more fully later in this document.

In addition to the general assumptions about system alignment, several key issues affect standards-assessment alignment. The ultimate quality of any aligned system rests on the simple fact that alignment means nothing if the content and performance standards are not carefully and thoughtfully developed and if assessments are not psychometrically sound. Shortcuts, while enticing, almost always produce less than desired outcomes. In essence, each component part in an aligned system must represent the best quality possible.

Elements of an Aligned System of Standards and Assessments

Three major elements of an aligned system of standards and assessments are the standards, the assessments, and the reporting of assessment results. The way in which the three elements are developed and linked will affect the quality of the alignment. The following sections summarize the development of these three component parts.

Standards Development

To specify and articulate the knowledge and skills that all students should have, *content standards* should represent rigorous academic content that is challenging to but attainable by all students; provide clear statements to teachers, students, and parents about expectations; represent the collective perspective of all educational stakeholders; be reached by consensus of the various stakeholders, including those involved in the development and review process and those who represent the larger public and business communities; set both long- and short-term goals for all schools to reach; promote good educational practice; and encourage students to actively evaluate and take responsibility for their own learning.[2]

Various processes and considerations can help ensure that stakeholders' opinions are incorporated into the development of content and performance standards. These include

- public forums for gathering information about what the public believes is important and what it expects students to learn in school;
- committees of educators having expertise in academic content, child learning and development, and pedagogy;
- advisory committees that can provide perspectives beyond those of educators; and
- pilot studies and other data-gathering activities that contribute to the setting of fair standards for describing student performances on standards-based assessments (e.g., proficient, advanced).

Performance standards should be interpretable in terms of the content standards on which they were developed; focused on learning and congruent with how learning actually occurs; grounded in student work but not tied to the status quo (the goal is to raise achievement levels, not to simply reflect "what is"); understandable and useful to teachers, students, and parents; and engage students in judging the quality of their own performances.[3]
Remember that performance standards should include multiple performance levels; descriptions of student performance at those predetermined levels; examples of the full range of student work at each level; and a series of methodically set cut scores used to categorize performance on assessments into each of the performance levels.

2 For additional information on the development of content standards, see Mitchell, R. (1996) *Front-end Alignment: Using Standards to Steer Educational Change*. Washington, DC: The Education Trust.
3 For additional information on the development of performance standards, see Hansche (1998).

Assessment Development

If an assessment is to achieve its purpose, validity, reliability, and fairness must be attended to during all phases of assessment development, from planning to form design. The quality of test development procedures affects the validity and reliability of the instruments. A detailed description of the test development process is beyond the scope of this document, but certain aspects relevant to alignment are highlighted.

The beginning point for any useful assessment instrument is clearly defining the domain of knowledge and skills to be assessed. If the assessments are built on agreed-on content and performance standards, a critical step in developing an aligned system will have been addressed.

Test blueprints or test specifications delineate the relative emphasis, or weight, for each content standard, usually recommended by various review committees. Test blueprints provide the framework for test form construction. The blueprints address the breadth of content coverage: Will each test form cover all the standards? Will each test form cover only a limited number of standards? Will each form simply sample from the domain of all standards? Test blueprints also may address how the difficulty of the various test forms will be handled. After a blueprint is determined, it becomes the basis for linking and equating test forms and data across test administrations from year to year.

The beginning stages of the development process usually includes creating *item specifications*. Item specifications should accurately reflect the content standards at each appropriate level and provide the contextual framework and measurement requirements for each item or task. Item specifications often answer questions such as: Are items to be open ended or selected response? Are they to be written at a level of simple recall and comprehension or at a higher level of thinking or at multiple cognitive levels? Is the content limited by including what may be assessed, or is it limited by describing what is not acceptable to assess? Are the item specifications matched to specific content standards and curriculum? Based on the specifications, individual items and tasks can and should be tied directly to content standards and curriculum. When properly written, item specifications can be used effectively to address depth and range of content coverage.

Moving from content standards to assessment items and tasks can have serious ramifications for alignment. Although creating item specifications from content standards and curriculum is often thought to be routine, the process is complex. A simple but classic example is a standard requiring that "students read a wide variety of literature genres," translated into the following test item:

> Read the (given) passage. This is an example of
>
> (A) poetry.
> (B) narrative prose.
> (C) expository writing.
> (D) technical writing.

This question must be posed about alignment: "Does this item measure the content as the standard was intended?" If the answer is "yes," the item is aligned; however, if the answer is "no," the assessment item is not aligned to the content standard. Often, answers to questions like this one are judgment calls. For that reason, standards developers, curriculum experts, and various stakeholders should be involved as participants and reviewers throughout the development process — most importantly as the content standards are translated into assessment specifications and ultimately into test items and tasks.

Good test development requires several levels of review during the development process. Reviews should include, at a minimum:

- *technical review* of items and tasks individually and as a set or pool of potential items prior to field testing (e.g., Is the item straightforward, unambiguous, and tied directly to a stated objective [both the assessment objective and content standard]? Does it reflect the intent of the standard? Is a range of performance levels addressed by the items and tasks in the pool?);
- *qualitative review* for potentially biasing elements prior to field testing (e.g., Is the item free of stereotypes, gender bias, racial bias, cultural bias?); and
- *statistical review* after field testing for psychometric qualities (e.g., p values, point biserials) and differential item functioning (bias based on race and ethnicity, gender, region, etc.).

Items that survive the review processes can be pooled for use on test forms. Once assembled according to the test blueprint, each test form requires review for breadth and depth of content coverage. It is only with careful attention to a sound development process that the assessments become valid and reliable indicators of student achievement required in an aligned system.

In addition to maintaining a sound test development process, some key qualities of an assessment instrument are prerequisite for establishing alignment. Validity and reliability are basic qualities that must be addressed continually during the test development process if an assessment instrument is to provide some form of accountability information and is to achieve its purposes.

The *validity* of an instrument refers to the degree to which the instrument measures what it is intended to measure (e.g., mathematics ability) and the degree to which measurement results can be used for their intended purposes (e.g., determination of proficiency). For example, a math test including word problems written in English may constitute more than a math test for students with limited English proficiency and results from their performance may not provide a clear determination of math proficiency.

Reliability is a statistical property referring to test score accuracy and consistency. The reliability of an instrument is a necessary component of validity because error in measurement undermines the intended use of measurement results. Reliability, however, is not sufficient to establish the validity of an instrument. To return to the example, the student with limited English proficiency may attain consistent math scores on repeated assessments (an indication of reliability), but if reading ability confounds math performance, the measure is invalid.

Reporting

As stated previously, aligning standards and assessments relies on broad system alignment and dynamic system elements. The elements are dependent on one another; changes in one part of the system will affect other parts. An aligned assessment functions appropriately only if it can be used to inform other aspects of the system, including standards. Because of this, appropriate reporting of assessment results and communication of student and school expectations are key elements of an aligned system. The U.S. Department of Education (USDOE) (1999) refers to this as "transparency" or the degree to which materials are readily available to teachers, students, and parents so that they can clearly see the relative weight of the standards in the assessments. The USDOE notes further that assessment results should be reported in ways that make it possible to target instructional improvement relative to the standards. Although this may not be practical at the individual student level, at least on the basis of the state assessment alone, school and district personnel should be able to detect how well their students as a group

are mastering the standards. Reported results should also be easy to understand (e.g., the percent of students who meet the standards).

Clear communication should not be restricted to the reporting of student assessment results or to the availability of standards documents to parents, teachers and school administrators; it should be more inclusive. The underlying policies should be in a form teachers and administrators can understand and use, and the public must feel confident that the system is truly aimed at helping students learn what is important and useful (Webb, 1999a). Moreover, a system should provide clear expectations to all stakeholders — this includes an indication of the purpose of testing prior to administration to students, clear definitions of the outcomes students are expected to demonstrate, the basis on which students are being judged, and exemplars and criteria of excellence related to outcomes so students know what is expected.

Results should be clear and understandable to all stakeholders, provided in various forms to stakeholders for their use in educational decision-making, and support collaboration among students, parents, and educators (Virginia Department of Education, 1993). As stated by Webb (1999a), standards and assessments must work together to provide consistent messages to teachers, administrators, and others about the goals of learning. For example, a system is not aligned, and students and teachers receive mixed messages, if the standards indicate that students should learn and contribute productively both as individuals and as members of groups, but no part of the assessment system (local or state) produces evidence of whether students are contributing productively as members of groups.

In developing standards-based assessment systems, progress in reaching standards must be communicated in meaningful ways for *all* students. Although all students are expected to master the standards, it is simply not realistic to expect that to happen immediately. There should be a sufficient number of performance levels to provide educators, students, and parents with clear information about the progress of children in moving toward higher levels of achievement, no matter how far they may be from the desired level of expectation (Carlson, 1996). For example, in addition to achievement-level reporting by categories (e.g., advanced, proficient, basic, below basic), reporting should be supplemented to show how close a particular student is to the next higher level or lower achievement-level boundary (National Research Council, 1999). For instance, was the student's score close to the upper boundary in the basic level and nearly in the proficient category? Or, did the student only just make it over the lower boundary into the basic category? Achievement-level descriptions should provide a clear picture across a broad range of performance levels with corresponding details related to each academic content area. The clear connection between reporting results and alignment of standards and assessments is supported in other work as well (Kentucky Department of Education, 1993; Pipho, 1997; Romberg and Wilson, 1995; Long and Benson, 1998).

Other Issues Related to Alignment

There are other issues to consider in developing or evaluating system alignment. Many of these issues are related to the purposes of assessments and standards and broad-based national considerations.

Federal Program Requirements

Certain federal requirements and regulations for Title I and other related programs should be considered when developing an aligned educational system. Elementary, secondary, and special education programs now rely on state standards-based assessment systems to evaluate the effectiveness of federal programs, instead of requiring separate tests for each federal program (Walkup, 1999). Accountability systems must identify low-performing schools and provide strategies for schoolwide improvement. The clear message

is that *all* students should be held to the same high standards. This includes students from low-income families, bilingual families, and migrant families, as well as most students with disabilities. Federal expectations for Title I are as follows:

- Academic standards must be rigorous — not minimum competencies.
- Assessments must be based on a state's standards by 2000–2001.
- Assessments must be fair, valid, reliable, and include all students.
- Assessment results should be reported for individual students, schools, and districts.
- Assessment results must be reported on the bases of gender, race and ethnicity, English proficiency, disability, migrant status, and low-income status.

Further, Title I legislation (Public Law 103-382, sec. 1111) specifies that performance standards must provide information for at least three levels: two levels of performance — proficient and advanced — that determine how well children are mastering the material in the state content standards, and a lower level of performance — partially proficient — to provide complete information about the progress of lower performing children toward achieving the proficient and advanced levels of performance.

Title I IASA legislation also makes it clear that state assessment systems must provide accurate inferences about the standards-based achievement of *all* students, including students with limited English proficiency and students with disabilities. In an aligned system, this means that the content standards, assessments, and performance standards must be free of bias for or against any group — they must be fair. It does *not* mean that standards and assessments should not be rigorous. In fact, the law requires rigor. Part of fairness is ensuring that *all* students receive adequate and appropriate opportunities to learn and demonstrate their learning.

One aspect of "fairness" is the establishment of *validity evidence for subpopulations* (Texas Department of Education, 1999). The guiding principle is that high academic standards, inferences of achievement, and access and opportunity for *all* students depend in part on assessments that are valid and reliable for all subpopulations of students. This requires a high degree of sensitivity regarding the impact of assessments on minority and special populations of students (see Kentucky Department of Education, 1993).

Standards of National and Professional Associations

One aspect of validity is the degree of association of one instrument with other instruments designed to measure similar constructs (criterion-related validity). For state tests, this often means examining the extent to which state assessments are related to national assessments such as the National Assessment of Educational Progress (NAEP), which is based on a national consensus about important content standards. These standards are developed by national professional organizations such as the National Council of Teachers of Mathematics (NCTM), the National Council of Teachers of English (NCTE), and the American Association for the Advancement of Science (AAAS). Considering national standards and assessments in the development process can enhance the credibility of state standards.

Norm-Referenced and Criterion-Referenced Tests

The decision to include nationally normed tests as part of an assessment system should be a function of purpose. If national comparisons are desired, a nationally normed test must be added to the assessment mix. Norm-referenced tests (NRTs) typically are built to match generic educational standards and curricula. Because it is based on generic content, a nationally normed test is unlikely to be strongly aligned to state standards and, thus, cannot provide an adequate measure of how well students or schools are achieving the state standards.

By contrast, criterion-referenced (CRT) or standards-based tests can provide this match. If the assessment is to establish the extent to which students are achieving and demonstrating proficiency in specific areas, criterion-referenced tests are paramount. The criterion-referenced instrument actually measures the degree of success or level of proficiency rather than distributing students along a "normal" curve. Thus, norm-referenced and criterion-referenced tests serve separate purposes. Each can be useful within a single assessment system; the extent to which data from either type of test are useful will depend on the purposes for their inclusion.

Multiple Measures

Without exception, more than one measure of performance is required to make valid and fair inferences regarding student and school achievement. Because no single form of assessment can be designed to serve all the purposes of large-scale assessment, it is necessary to use a variety of assessment techniques appropriate to different purposes (NCTM, 1989; Costa, 1989). Multiple measures can mean using different measures of a particular construct or aspects of the construct, taken at one time; and it can mean administration of similar measures at different times. For example, accurate assessment of students' writing achievement may require administering several writing prompts or a writing prompt in combination with an editing exercise. To assess students' progress, different writing prompts or editing exercises may be administered over time.

SUMMARY

An aligned system of state standards and assessments exists within a larger educational framework. The influence of curriculum and instruction, the clarity of state expectations, and a shared philosophy between state and local educational agencies are but a few of the assumptions that underscore the development and evaluation of system alignment. Among other things, best practices as they relate to standards and assessment development, federal legislation, and the need for multiple measures constitute critical considerations when studying alignment.

Developing and evaluating an aligned system of standards and assessment is discussed next, along with a discussion of structures and mechanisms of support that allow for maintenance of an aligned system.

II. DEVELOPING AND EVALUATING SYSTEM ALIGNMENT

It is common sense to align an assessment system with content standards to provide useful information about student and school achievement. Students should be required to participate in state assessments only if and when it is clear exactly what information is being sought and how that information will be used. Developing and evaluating an aligned system involves addressing complex issues and careful planning to achieve that near perfect state of alignment.

GENERAL ORGANIZING PRINCIPLES

There is a relative paucity of research addressing the thorny issue of alignment, specifically alignment between assessments and standards. However, the available and existing work focused on this issue points to five interrelated dimensions that should be considered when determining the extent of alignment within an educational system.

Content Match

Alignment depends largely on a close match between content standards and assessment content. Content match may be considered a necessary condition for an aligned system of assessments, but alone it is not sufficient to produce a high degree of alignment. The USDOE (1999) indicates that all standards must be assessed, referring to the "comprehensiveness" of the assessment system. This point is reiterated in the notions of "range of knowledge correspondence" and "categorical congruence," meaning that both standards and assessments cover a comparable span of topics and ideas within categories and at the specified level of detail (Webb, 1999a). NCTM (1989) indicates that the set of tasks on the assessment instrument must reflect the goals, objectives, and breadth of topics specified in the curriculum and that, ideally, all topics in the curriculum should be assessed (see also Virginia Department of Education, 1993). The breadth of coverage is not limited to academic standards but should include a broader vision if dispositional characteristics are noted within the standards. Webb (1999a) terms this aspect of alignment "dispositional consonance."

It is improbable that a single assessment instrument will provide the breadth of coverage necessary for an aligned system. This, of course, depends on the number of standards and their specificity. To effectively cover the breadth of the standards without overburdening students, sampling approaches may be required (e.g., sampling students or using multiple test forms). Moreover, the USDOE (1999) indicates that local assessments may be needed to supplement state assessments and suggests further that an assessment system designed to identify school-level performance should not be based on a single instrument.

Depth Match

Webb (1999a) indicates that standards and assessments are aligned if they reflect similar requirements on the number of dimensions covered. This may include the level of cognitive complexity of the information students are expected to know, how well they should be able to transfer this knowledge to different contexts, and how much prerequisite knowledge they must have to grasp more sophisticated ideas. To use an example from Webb (1997b), "the Curriculum and Evaluation Standards for School Mathematics published by the National Council of Teachers of Mathematics (1989) states that students in grades 9

through 12 should study data analysis and statistics so that all students can 'design a statistical experiment to study a problem, conduct the experiment, and interpret and communicate the outcomes.' An assessment system requiring students only to interpret an existing set of data would not be aligned with the depth of knowledge specified in this standard" (pg. 5). The USDOE (1999) similarly suggests that assessments should be at the same levels of breadth and depth of content and skill coverage as are the content standards (see also Walkup, 1999).

Moreover, Webb makes reference to cognitive complexity in evaluating alignment, citing a strong research base (e.g., Romberg and Carpenter, 1986; Stein, Grover and Henningsen, 1996). These researchers contend that research addressing how students develop knowledge within content areas should be considered in evaluating the cognitive soundness of an assessment system and that this may be revealed in the articulation and assessment of standards across grades and ages. Aligned standards and assessments are complementary in their representation of the underlying structure of knowledge students need to develop and how their instructional experiences should be organized.

Emphasis

The degree of alignment also depends on the extent to which the assessment's relative emphases on various topics and processes reflect the curricular and standards emphases. An assessment instrument that contains many computational items and relatively few problem-solving tasks, for example, is poorly aligned with a standard that stresses problem solving and reasoning. Similarly, an assessment instrument highly aligned with a standard that emphasizes the integration of mathematical knowledge must contain tasks that require such integration (NCTM, 1989). Webb (1999a) supports this claim, indicating that standards and assessments should embody similar requirements for the ways students should draw connections among ideas. Webb (1997b) indicates that a "balance of representation" is needed in which similar emphasis is given to different content topics. The standards and assessments should give comparable emphasis to what students are expected to know and be able to do, and in what contexts they are expected to demonstrate their proficiency. Drawing again on the examples provided by Webb (1997b), the National Science Education Standards (1996) emphasize different skills at different grade levels. Students in K–4 are expected to focus on developing observation and description skills while students in higher grades are expected to work on constructing models that explain visual and physical relationships. An aligned assessment system would need to reflect a similar shift in emphasis and include enough different tasks to reflect the same priorities and intentions as the standards.

A related consideration is the degree to which the assessment is intended to measure the knowledge and skills specific to a set of single grade-level standards or to measure cumulative knowledge and skills spanning several grades. For example, an assessment used in grade 8 might include items and tasks designed to assess cumulative knowledge of lower grade expectations, but some emphasis must be placed on eighth-grade expectations. The purpose of the assessment and the structure of the knowledge and skills in the content area will contribute to decisions about emphasis.

Performance Match

Students should be assessed in a manner that reflects the nature of performance described in both the content and performance standards (USDOE, 1999; see Hansche, 1998, for a detailed discussion of the role of performance standards). As noted previously, Hansche (1998) defines a performance standard as a system including performance levels (labels), performance descriptors, exemplars of student work, and cut scores. Therefore, providing an aligned assessment system requires a match relative to these elements. For example, if cut scores for levels such as "proficient" or "above standard" are prescribed as part of a performance standard, an assessment instrument or system should provide commensurate scoring information to determine student performance (Walkup, 1999).

Performance match is perhaps more difficult to accomplish when considering performance descriptors that may or may not include examples of student work. Performance descriptors may not lend themselves easily to traditional large-scale assessments, especially off-the-shelf products, but may imply a specific assessment instrument, task, or item type (NRT, CRT, multiple-choice, constructed-response, performance-based).

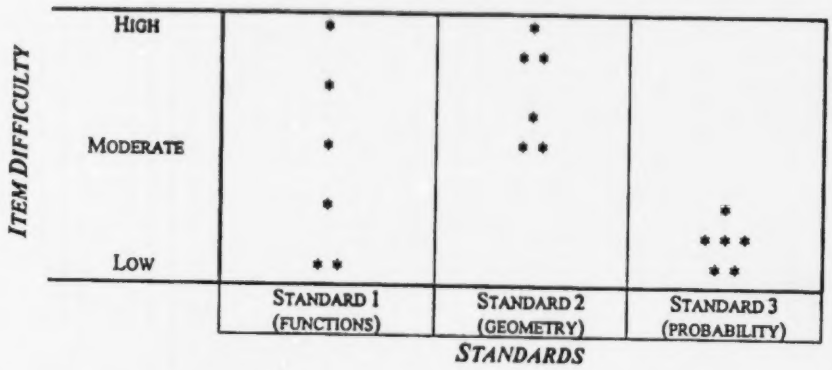
Performance descriptions may reflect instructional approaches and activities such as the use of manipulatives, calculators, and computers. When these materials are used during instruction, they should be available during assessment, as long as their use is consistent with the purpose of the assessment. For example, if students routinely use calculators for solving problems in class, they should also be able to use calculators during assessments of their problem-solving abilities. Similarly, if students' understandings are closely related to the use of physical materials, they should be allowed to use these materials to demonstrate their knowledge during the assessment (NCTM). Webb (1999a) echoes this element of match especially as it relates to technology, indicating that standards and assessments should send students consistent messages about technology and how it is related to what they are expected to learn. If standards indicate that students should use calculators or computers, instruction should provide adequate opportunity for students to use them and assessments should do the same. Unfortunately, traditional forms of student assessment and the constraints imposed by limits on time and other resources may result in an inordinate emphasis on the superficial acquisition of skills and facts. The consequence would of course be an assessment system with only minimal alignment to performance standards.

Accessibility

Thus far, all dimensions for evaluation of system alignment reflect the match of standards and alignment assuming applicability for *all* students. It is necessary to make this assumption explicit. When the expectation is that all students achieve high standards (as in Title I IASA legislation), aligned assessments must give every student a reasonable opportunity to demonstrate attainment of the standards. An aligned system will demand equally high learning standards for all students, while providing fair means for all students to achieve the performance standard (Walkup, 1999). Because student performance on assessments depends on a number of factors other than level of knowledge, assessments will be more equitable if multiple measures are used (Webb, 1999a; see also Virginia Department of Education, 1993). In working with assessment contractors, effective monitoring is needed to protect against hierarchical implications in skills and knowledge that are inappropriate or unreasonable in test construction and item development (National Research Council, 1999).

An aspect of accessibility is the necessity to include, for each measured standard, assessment items and tasks that vary in terms of item difficulty, spanning different levels of achievement. For example, Figure 1 illustrates how three math standards are assessed on a hypothetical test (for simplicity, the test is depicted as measuring only three standards with six items per standard). Accessibility for students with different achievement levels is only evidenced for standard 1 (functions). Although six items measure standard 2, the items are all relatively difficult. Students with lower levels of geometry knowledge and skills will be unable to demonstrate any knowledge with respect to the standard. Conversely, standard 3 (probability) is measured by only relatively easy items. Higher achieving students will not be able to demonstrate their full range of knowledge and skill on this standard. As a whole, student scores on the assessment would not accurately reflect achievement in geometry and probability.

Figure 1. Item Difficulty and Standards Measurement.



Another way to depict this accessibility issue is through linking content standards, performance standards, and test items as in the following display. This sort of display could be repeated for all content standards.

Standard A	Exceeds Standard	⇨	Test Items
	Meets Standard	⇨	Test Items
	Approaches Standard	⇨	Test Items

Both of the above examples assume items with a single difficulty level. One solution to making assessments more accessible is to include items and sets of items that are designed to elicit responses from students representing a wide range of knowledge and skill. A familiar example is a writing prompt with a multilevel scoring rubric. The prompt and scoring rules provide information about students' writing skills on a continuum of performance. A major benefit of using such constructed-response items is their potential for measuring knowledge and skills across a range of achievement.

Another aspect of accessibility that supports alignment for all students involves bias. If in place, best practices associated with test development can help solve this potential problem. Monitoring for bias, including both qualitative and statistical bias review, can help promote the alignment of an assessment system (Virginia Department of Education, 1993).

All students, including students with disabilities and English language learners, must have the opportunity to demonstrate achievement and attainment of standards. Assessments are the primary vehicles used to gauge achievement and skill attainment; therefore, a variety of strategies will be needed to make certain that all students participate in the assessment system. These strategies may include accommodations, alternate assessments, assessments in students' primary languages, and linguistically simplified assessments. If students are exempted from the required assessments, descriptions of the exemption criteria should be provided, as should the procedures used to determine what modifications to assessment procedures are needed. Procedures for documenting the assessment of all students, including auditing and record-keeping, should also be explained and be made publicly accessible (USDOE, 1999; see also Maryland State Department of Education, 1996).

METHODOLOGY

For nearly all assessment purposes, an underlying assumption is that the data gathered through assessment will be used to make decisions regarding the educational process. If school accountability is the primary purpose, data gathered through assessment will be used to judge schools and for planning purposes to overcome deficiencies. If passing an exit examination is a requirement for high school graduation, the examination will be scored to determine acceptable levels of proficiency and results will undoubtedly be used for purposes of remediation. Because there are important reasons and purposes for use of the data gathered through assessment, the alignment between state assessments and standards is of utmost importance.

As a consequence, evaluation of alignment should not be entered into lightly but should follow systematic procedures (both qualitative and quantitative) that allow conclusions regarding alignment. In his study of alignment practices, Webb (1999a; see also Webb 1997b) found that states have approached alignment in three main ways:

- In a serial manner, whereby one party works on content standards, another entity works on assessments, and, at some point, the "pieces" are put together;
- Via a "checklist" procedure whereby, a test developer checks off the standards as they are covered in the assessments; and
- Through an external evaluation, whereby a third party evaluates the degree of match between the standards and the assessments.

This is probably not an optimal list of evaluation and development methods. Given the desire to connect standards and assessments, it would be difficult to achieve a high level of alignment if the two pieces were developed and evaluated in relative isolation. It would probably be equally difficult for a test publisher to provide an objective evaluation of an instrument it has developed. Although a third-party evaluation may be a plausible alternative, affording a high level of objectivity, nuances and intentions embedded within the development of both standards and assessments may be critical to the evaluation process. Unless this information is explicitly shared, the evaluation provided by a third party may be of limited value.

There are numerous ways that an evaluation of the alignment between assessment and standards could be conducted. Several examples are provided here to serve as models; this is not an exhaustive list of methods and inclusion here is not intended to constitute a value judgment. The purpose is to provide examples of how researchers have approached these important alignment issues using *systematic* procedures.

1. Webb (1999a; 1999b) conducted a study of alignment between standards and assessments in four states using methodology developed partly at the Institute for the Analysis of Alignment Criteria. Another purpose of this study was to develop methodological criteria that could be systematically applied in an alignment evaluation.

   In this study, the match of standards and assessment content in terms of breadth and depth of knowledge was reviewed. Reviewers were provided with several specific levels of criteria to judge depth of knowledge, including recall, skill/concept, strategic thinking, and extended thinking. Reviewers first applied these criteria to content standards in order to estimate the depth of knowledge necessary to attain a standard. A review of each assessment task followed, using the same criteria for judging depth of knowledge.

Three to five individuals reviewed the match for a specific content area and for a specific grade. There is no indication made about the necessary number of reviewers, but the use of multiple reviewers has several advantages, including providing a greater level of objectivity and allowing for an analysis of the agreement between reviewers (interrater reliability).

In a spreadsheet format, reviewers coded their judgments first for each content standard/objective and, in adjoining columns, coded whether or not an assessment item or task matched a content standard and, if so, at what depth-of-knowledge level. Following this procedure, independent reviewers selected a sample of judgments for comparison with other reviewers. This process allows evaluation of interrater reliability and also a point of feedback that can be used to reconsider judgments.

Generally speaking, a standard was judged to have an acceptable degree of alignment if six or more items measured the standard and 50 percent or more of the matching items were at or above the necessary depth of knowledge. Webb (1999b) indicates that the number of items necessary for an acceptable match is somewhat arbitrary but should be determined based on acceptable levels of reliability. Using the six-item criteria, the four states studied were found to have relatively low degrees of standard-assessment alignment.

2. Wixson (1999) borrowed the methodology developed by Webb and Blank in a study of alignment within four states. Wixson emphasized the need to consider the state history and experience involving standards and assessment development. Elements such as test blueprints and item specifications are essential to providing insight into the assessment development process.

   Alignment was judged in terms of overall coverage and depth of match. Criteria for an acceptable level of match were more liberal than those used by Webb. In this study, a single test item matched to a given standard was judged to provide an acceptable level of alignment. Again, no empirical base was provided as a justification for the one-item criteria.

   Two of the four states were judged to have relatively high levels of alignment, one was judged to have a moderate level of alignment, and one was judged to have a low level of alignment. Knowledge of the history of standards and assessment development provided further interpretive information. In one instance, the development of standards was driven in large part from the current state assessment. Therefore, relatively high alignment was a consequence. In a second case, standards were presented in such a broad-based fashion that alignment in terms of breadth of coverage was easy to achieve.

   With all states, the depth of knowledge match was judged to be relatively low. Taking into consideration dimensions of alignment beyond simple content match provides a more rounded evaluation of alignment.

3. Schmidt (1999) has suggested a methodological approach to the study of alignment similar to that of Webb (1999a) and Wixson (1999), although it is different in one significant respect. As in the other approaches, both standards and assessments are evaluated, but in the Schmidt approach, coders or raters are responsible for coding either some aspect of an assessment or standards document but are not responsible for matching the two halves (blind review).

   In essence, both sets of documents are coded independently in terms of a defined, exhaustive set of content specifications, and the actual match becomes a statistical comparison of coded objectives. Schmidt adopts this strategy in an attempt to diminish the introduction of bias that can creep in when it is a coder's job to identify a match between a known standard and an assessment item.

   As in the Webb methodology, multiple coders are used and information is coded on multiple dimensions of alignment. In a study of 20 states, Schmidt found relatively low levels of alignment.

4. Romberg and Wilson (1995) reviewed two studies they and their colleagues conducted addressing the alignment of six standardized mathematics tests to the grades 5 to 8 NCTM standards. To demonstrate the extent to which the tests reflect the NCTM standards, items were classified by multiple raters using three sets of criteria, including content, process, and level of response.

   The content areas included numbers and number relations, number systems and number theory, algebra, probability or statistics, geometry, and measurement. The process categories specified in the standards were problem-solving, communication, reasoning, connections, computation and estimation, and patterns and functions. Two levels of response were considered: concepts or procedures (according to whether the response required conceptual or procedural knowledge).

   Alignment of the six standardized assessments to the NCTM standards was judged by the researchers to be poor.

5. A different type of alignment study is described by Sanford and Fabrizio (1999). This involved evaluation of alignment between the North Carolina End-of-Grade Test of Mathematics at grade 8 and the National Assessment of Educational Progress (Mathematics, Grade 8) along three dimensions: technical, content, and cognitive demand.

   The technical characteristics examined by expert panels were item format, number of items, distribution of items by content area, administration time, and items administered at multiple grade levels. The content characteristics examined by the expert panels were data analysis, statistics, and probability; measurement; algebra and functions; geometry and spatial sense; and number sense, properties, and operations. Conceptual understanding, procedural knowledge, and problem-solving were considered in judging high vs. low cognitive demand.

   A conclusion was that the content and depth of coverage offered by the separate assessments did not align well. Among the reasons for this lack of match were substantive test specification and test framework ("blueprint") differences.

These five studies highlight possible approaches that can be used to evaluate standards-assessment alignment. There are commonalties that can be identified among the approaches. In each of the approaches, multiple individuals were responsible for judging alignment, and reviewers were responsible for judging alignment using a set of predetermined criteria. The specific criteria used overlapped from study to study, but different questions were asked and the level of quantification varied from study to study. Perhaps the overriding commonality and the intended point of this discussion is that a systematic approach should be used to evaluate alignment.

The described methods allow an evaluation of alignment as it relates to content, depth, and performance. It seems reasonable that these methods could be easily adapted to allow for an evaluation of alignment as it pertains to emphasis. For example, the approach taken by Webb is also used to rate the relative emphasis of particular goals and objectives within standards.

By contrast, the dimension of accessibility may require additional evaluation methods. A qualitative review of standards may provide insight into its applicability for all students, and technical aspects of assessments (e.g., provision of accommodations, bias review committees) may also speak directly to assessment accessibility. Post hoc statistical review is necessary to identify the extent to which items both discriminate between high- and low-achieving students and allow students from the entire spectrum to express skill. Post hoc statistical review is also part of the ongoing development of an assessment system that is free of bias toward specific groups of students. This sort of review not only highlights assessment inadequacies but can shed light on the accessibility of a standard to all students.

Reporting as an aspect of alignment may not lend itself to a direct quantitative evaluation, but it is worthy of consideration. Reviewing reporting documents can provide insight into how information provided by the state is translated into educational practice. Focus groups involving teachers and parents may be particularly useful in identifying the usability of reported information. Survey and observational methods also could be employed to investigate the impact of reporting strategies as they relate to instructional practice.

SUMMARY

The purpose of this discussion was to provide insight into available methodological strategies for studying alignment. An actual review of alignment was not intended, but the fact that each of the described studies concluded that standards-assessment alignment, by and large, was lacking cannot be escaped. This should not be considered an indictment of state assessment systems; however, the information provided by these studies is valuable. Knowledge of specific weak points within an assessment system provides insight into how to correct or plan for the future. For example, Webb (1999b) suggests that one benefit of alignment is reduction in instructional redundancy. Further, knowledge of areas of alignment strength can provide the basis for an analysis of the effectiveness of instructional strategies.

There are other benefits to studying alignment. As stated by Sanford and Fabrizio (1999), the North Carolina evaluation resulted in a clearer understanding of the assessment instruments. A better understanding of contributions from both instruments, given their separate purposes, was gained. Moreover, at first glance the lack of match between the North Carolina and NAEP assessments seems troublesome. However, if both instruments, serving separate purposes, independently add to the match between the assessment system and the state standards, the lack of instrument match may be considered a positive factor. The Sanford and Fabrizio study also illustrates the various layers of the educational system that can be evaluated and that pertain to systems alignment.

Although the focus of this paper has been standards-assessment alignment, recognition of a broader educational scope (systemwide alignment) was also provided. The Sanford and Fabrizio study points to yet another layer of alignment (alignment within an assessment system) embedded within our general focus.

III. Supporting Mechanisms and Structures of an Aligned System Focused on Student Learning

To restate a general consideration, the alignment between standards and assessments is a process embedded within a broad educational context and should be viewed as part of comprehensive system development. Comprehensive education systems require that all parts be linked and work together as a whole (Hansche, 1998).

As an aligned system is constructed, attention shifts to maintaining the system. Perhaps the primary mechanism of maintenance and continued development and revision is to use data garnered through the state assessment system. The use of assessment data may manifest itself in a variety of ways, including changes in instructional practices and strategies, focus of instruction on areas of greatest need, and knowledge of those aspects of standards that must be taught and assessed solely within the classroom. State assessment data also can be used in revising standards and in planning the development of supplemental assessment pieces. In short, data gathered through state assessments provide feedback for systems maintenance and again focus attention squarely on the reporting aspect of the assessment system.

Many other components contribute to the effectiveness of an educational system, including but not limited to standards, performance and assessment, learning readiness, resources, parent involvement, safety and discipline, technology, professional development, school accountability and school autonomy. In addition to the alignment of the critical elements of standards and assessments discussed so far, several policy issues and related educational components should be considered. These other mechanisms and structures support and help maintain an aligned system after it has been developed. The following are selected examples of some of these support mechanisms.

Accountability: Accountability is the application of consequences to assessment results. Accountability can occur at the state, district, school, classroom, teacher, or student level. Linn (1998) indicates that history has shown that testing is a popular instrument of accountability and reform for several reasons, including that tests are relatively inexpensive, testing changes can be implemented relatively quickly, test results are visible and draw media attention, and testing can create other changes that would be difficult to legislate (e.g., curriculum). Moreover, assessments can provide valid and reliable information about student performance. Accountability is fair and defensible to the extent that the inferences made on the basis of assessment results are accurate. For assessment results to yield fair and accurate inferences, the assessments must accurately reflect the knowledge, skills, and cognitive demands of the standards.

Teacher involvement and professional development: Instruction aligned to standards is critical for students to attain high standards, as measured through aligned assessments. Teachers and administrators must provide leadership and expertise in standards and assessment development and alignment processes. Teachers should be involved in improving the rigor of classroom teaching, strengthening the curriculum, selecting appropriate professional development, determining how to use standards and new assessments, and strengthening accountability systems at all levels. Without adequate professional development opportunities, it will be difficult, if not impossible, for teachers and administrators to meet the demands for an aligned curriculum,

including classroom assessments, to provide continuous information regarding student progress relative to the standards.

Policy development: Clear policies must support sound practices in the development and adoption of standards and assessments that are both rigorous and fair to *all* students. Policies must also support procedures that ensure alignment and ongoing professional development. Similarly, the fiscal impact of these social policies is substantial; therefore, for standards-based reform to be effective, adequate financial support is required to get the job done (Pipho, 1997). The principles undergirding these policy components must work together. They imply a strategy of complementary actions, not a menu from which some choices can be taken and others overlooked.

Textbook adoption and use: In a standards-based environment, textbooks are critical tools in supporting student learning. Thus, the criteria for selection and the means for matching standards-based criteria to textbooks are matters of great importance for those who wish to improve teaching and learning. Textbooks, however, as a single source, cannot be expected to align perfectly with the breadth and depth of standards or the curriculum. This implies that funding must be allocated to develop and obtain instructional materials other than textbooks and that teachers will need opportunities to develop new strategies for using these alternate materials. There is also a need for qualitative evaluations of all instructional materials to ensure alignment and a high level of coherence with state- and district-adopted content standards and related assessments.

K–16 connections: For the most part, K–12 and postsecondary education systems are not coordinated (Kirst, 1998). States administer tests designed to assess student achievement on state-adopted standards. These assessments may include a variety of question formats such as multiple choice, writing, portfolios, projects, and other performances. At the collegiate level, admission policies rely heavily on college entrance examination scores designed to predict college success and a smorgasbord of placement test scores. The current array of policies and testing practices governing high school completion and college entrance send vague and confusing signals to students about what is needed to succeed. Although state standards and assessments that reflect alignment with national standards can increase the likelihood that students' learning is not parochial, there is a disconnect between what students must learn to fulfill high school graduation requirements and college admission requirements.

IV. CONCLUSION

Systems alignment might be considered the foundation for standards-based reform. Alignment involves a complex set of issues that surface at different levels of the educational system. The alignment of standards and assessments depends on state and systemwide alignment. For example, Linn (1998) suggests that academic standards at the national, state, and district levels often are inconsistent. As discussed by Bailey and Ross (1998), lack of alignment between state and local educational agencies may render alignment between state standards and assessments powerless.

Systems alignment, and the alignment of standards with assessments in particular, is a dynamic and cyclical process. It requires constant attention throughout development, evaluation, and re-evaluation. Standards can provide a framework from which to develop assessments, which may in turn provide information regarding the attainment of standards. Assessment results provide the opportunity for data-driven decision-making. The decisions made may involve instructional or curricular change. Assessment data may also affect a decision to modify expectations for student knowledge and skill development. In short, alignment can improve teaching and learning relative to standards (Long and Benson, 1998).

Several issues related to standards-assessments alignment were considered within this document. The technical quality of both standards and assessment will undoubtedly affect the quality and degree of alignment. The purposes of an assessment system will determine the degree of alignment necessary. For example, accountability systems probably require more than one assessment instrument (Linn, 1998). Moreover, certain federal requirements necessitate and guide the development of aligned standards-based systems.

There are several aspects of alignment between standards and assessments that should be considered during developmental and evaluation stages. These include content match, depth match, emphasis match, performance match, reporting systems, and accessibility to all students. It has been suggested that a variety of approaches is available to evaluate the level of system alignment and that while a broad spectrum of choices are available, a systematic approach to the study of alignment should be undertaken.

The alignment of standards and assessments depends on system wide alignment but also has implications for related educational considerations. Professional development, policymaking, and textbook adoption are but a few educational issues that will be affected by aligned educational systems.

In conclusion, systems alignment is a dynamic process. It is important to study alignment locally, at the state level, and at the national level. Alignment should not be seen as an all-or-none proposition but as one existing on a continuum from less to more. The goal should be to increase alignment so that valid and effective data-driven decision-making can be accomplished. A state that identifies a low level of alignment should work actively to improve that situation, knowing that alignment can be accomplished through the development process. It is better to know where weaknesses exist so that a course for improvement can be plotted than to avoid the issue or to assume that one is where one wants to be.

An aligned system of standards and assessments alone will not provide the increase in the quality of student learning that is the goal of standards-based education. Alignment among all components of the educational system, with logical connections among the components, the provision of appropriate resources to teachers and students, the involvement of parents and the community, and connections between K–12 and higher education are just some of the additional conditions needed.

APPENDIX: STANDARDS-ASSESSMENT ALIGNMENT CHECKLIST

This list describes key considerations in evaluating the degree to which an existing assessment system is aligned with content and performance standards; it is derived from research studies on alignment and from the experiences of state assessment staff in developing and reviewing assessments. Although the list is written from the perspective of an ex post facto evaluation of system alignment, it also can be used in developing an aligned assessment system.

As evidenced by recent legislation and research, the concept of alignment is ever-broadening. Therefore, although as many aspects of alignment as possible are incorporated the list is likely to be incomplete. An attempt was also made here to make the list broad enough to apply to the many assessment programs states and others have developed, while providing enough specific information to be useful.

PARTICIPANTS, TRAINING, AND MATERIALS

Webb (1999a) suggests that both content experts and people knowledgeable about a state's standards and assessments serve on panels that review systems of alignment. The latter panelists can explain how the standards and assessments are intended to be applied and can clarify other issues that arise. In addition, reviewers should be trained in the review process and monitored throughout the process to ensure that they are applying the review criteria appropriately (Webb, 1999a).

Participants

States should consider including panel members with expertise in

- the content of the standards and assessments;
- the students to whom the standards and assessments apply;
- the development and intended use of the content standards, performance standards, and assessment system;
- curriculum and instruction; and
- educational measurement.

Training

Panelists should be familiar with these topics (depending upon the composition of the panel, some of these topics may not need to be covered before review):

- content standards;
- performance standards;
- use and purposes of the assessments;
- student population to which the standards and assessments apply; and
- review process (training should include practice in the process).

Materials for Review

Panelists should have access to these materials during review:

- content and performance standards;
- assessment blueprints and item and task specifications;
- answer keys, scoring rubrics, and scoring guides;
- assessments;
- student response information (including sample responses for open-ended items and item and task statistics); and
- score reports at various levels of reporting (e.g., student, district, state).

REVIEW CATEGORIES

Alignment is defined here as the degree to which assessments yield results that provide accurate information about student performance regarding academic content standards at the desired level of detail, to meet the purposes of the assessment system.[4] For example, a state may want its assessment to produce information about the overall mathematics proficiency of its fourth grade students referenced to five defined performance levels; another state might want both information about overall proficiency and information about student achievement in particular areas, such as concepts and operations, geometry, functions, measurement, and probability.

To satisfy this definition, the assessment must adequately cover the content standards with the appropriate depth, reflect the emphasis of the content standards, provide scores that cover the range of performance standards, allow all students an opportunity to demonstrate their proficiency, and be reported in a manner that clearly conveys student proficiency as it relates to the content standards. The categories for review that follow are arranged according to these characteristics of the assessment, which were described earlier in this document.

Although it may be tempting to develop scoring rules for each characteristic of an aligned assessment (e.g., 95 percent of the items and tasks must relate directly to a content standard), evaluating the quality of alignment requires a holistic judgment. The purposes of the assessments and standards, their use in guiding instruction and decision-making, and other contextual information must be considered in judging whether the degree of alignment is sufficient.

Content Match

There is some controversy over the degree to which an assessment must match a set of content standards. Should each standard be represented by one or more items and tasks on the assessment? Can the assessment exclude certain groups of standards? Is domain sampling allowable in an "aligned" assessment? These questions cannot be answered without consideration of the nature of a state's content standards. Some states have broad standards, with fewer than 10 per content area in any grade level. Other states have more detailed standards, in some cases more than 30 per content area. State content standards also vary in grade level match. Some states develop content standards that are intended to be covered in a range of grades (e.g., grades 3 to 5); others have sets of content standards for each grade. As defined above, an aligned assessment yields results that provide accurate information about student performance at the desired level of detail. For states with broadly defined content standards, it is unlikely

4 For clarity, we refer to an assessment rather than an assessment system. However, these criteria can be and are intended to be applied to a system (e.g., of state and local assessment instruments) as well as to a single instrument. We refer to student scores as the level of reporting; these criteria can be applied to any level of reporting (e.g., school, district).

30

that an assessment that omits a content standard will yield results allowing inferences to be made about student proficiency in a content area. For states with detailed content standards, sampling within chunks of content (e.g., from the standards in the "probability" domain) may provide data that can be used to make inferences at the level desired. In either case, an aligned assessment does not omit any category of learning that is important to the content area as a whole, as defined by the standards.

The correspondence between content standards and assessment tasks may not be one to one, especially with assessments that include complex items and tasks,. For example, an "enhanced" multiple-choice item or a constructed-response task may assess more than one content standard, particularly if the standards are detailed. Conversely, it may take a number of multiple-choice items or short-answer tasks to adequately assess a broadly stated content standard.

Considerations in evaluating content match include:

Assessments are designed to match the content standards.

- Item and task specifications (for selected-response items and their options and for constructed-response and performance items and their scoring rubrics) specify the ways in which each standard will be assessed.
- The assessment blueprint describes how each content standard will be assessed, with appropriate item/task formats for each aspect of the standards.
- The blueprint specifies the proportions of the assessment that will cover each content standard.
- If domain sampling is used, the blueprint includes methods for ensuring that each domain is adequately covered.

All items and tasks are related to the content standards.

- Each item and task on the assessment measures part or all of one or more content standards.
- For selected-response items, incorrect options are related to inadequate or incomplete knowledge in the standard(s) assessed.
- For constructed-response and performance items, all criteria in the scoring rubrics are related to the standard(s) assessed.
- The items and tasks do not require students to use knowledge and skills irrelevant to the content standard(s) assessed. (For example, when skills such as reading are necessary for solving mathematics tasks, care should be taken that problems are worded clearly and simply, and appropriate accommodations are available for students who require assistance in reading.).
- The contexts (e.g., story problems, graphics, texts) in which items/tasks are set are appropriate to the content standard(s) assessed.

The assessment fully covers the content standards.

- All content standards or all important domains of the content area are measured by the assessment.
- Each content standard (or domain) is measured using an appropriate mix of item and task formats.

Depth Match

Most content standards imply a degree of cognitive complexity or level of difficulty of the concepts and processes contained in the standards. Webb (1999a) explains this relationship as "*...what is elicited from the students on the assessments is as demanding cognitively as what students are expected to know and do as stated in the standards*" (page 7; italics in the original). To evaluate depth match between content standards and assessments, states should first categorize the complexity of each content standard.

Considerations in evaluating depth match include

- Item and task specifications indicate the depth at which knowledge and skills should be measured.
- Each item and task elicits responses reflecting the depth of knowledge and skills in the content standard(s) it measures.
- Each item and task uses an appropriate format for the depth of knowledge and skills in the content standard(s) it measures.
- As a whole, the assessment reflects the range of depth of knowledge and skills implied by the set of content standards.
- Statistical item and task analyses indicate that items and tasks are at a level of difficulty commensurate with the content standard(s) measured

Emphasis

An aligned assessment should cover the knowledge and skills in the content standards with the same degree of emphasis as the standards. In general, this means that the *score* the student receives on the assessment should be based on the same balance of knowledge and skills as implied by the standards. In some cases, the raw number of items related to each aspect of the content area will indicate the relative emphasis. However, in assessments with items and tasks that can receive various numbers of score points (e.g., an assessment consisting of selected response items scored right/wrong and constructed-response items scored using either 3- or 4-point rubrics), the proportion of the total score should be considered in evaluating emphasis.

Considerations in evaluating emphasis match include:

- The items and tasks as a whole measure knowledge and skills in a manner that reflects the emphasis of knowledge and skills in the content standards.
- The formats used to measure different standards reflect the emphasis of types of knowledge and skills in the content standards.

Performance Match

An aligned assessment yields results that can be mapped onto the performance levels it is intended to measure. Performance match depends on both the difficulty and the content of the items and tasks. The content of the assessment must match the knowledge and skills described in performance descriptors for each level.[5] For example, performance descriptors may describe (in part) a partially proficient student as able to solve simple algebraic equations, a proficient student as able to evaluate algebraic equations, and an advanced student as able to use algebraic expressions in solving problems. The assessment should

5 Mills and Jaeger (1998) conducted a study in which they revised performance descriptors to match test content and found that where cut scores are set may depend in part on the match between test content and descriptors of achievement levels.

contain items and tasks that allow students at each level to demonstrate their knowledge and skills. This, of course, is also related to the issue of accessibility.

Considerations in evaluating performance match include the following:

- The assessment blueprint specifies how the entire range of performance descriptors will be measured by the assessment.
- Item specifications are referenced to the levels of knowledge and skills in the performance descriptors.
- The assessment as a whole covers knowledge and skills at each defined performance level.
- Each aspect of the performance descriptors is covered by one or more items and tasks.
- Score reports and statistical item and task analyses indicate that students at all performance levels have the opportunity to demonstrate their knowledge and skills

Accessibility

An aligned assessment provides all students in the system with the opportunity to demonstrate proficiency in content standards. The assessment should allow students who have learned the content in a variety of ways and mastered the content to varying degrees, students with disabilities, and students who are English-language learners to demonstrate content knowledge and skill. The assessment should be free from bias.

Considerations in evaluating accessibility include:

- Groups of selected-response items cover a variety of ways of expressing knowledge and skills related to the content standard(s).
- Constructed-response and performance tasks allow a range of responses to be referenced to each point in their scoring rubrics.
- Sample student responses for constructed-response and performance tasks contain a full range of response types and levels.
- Accommodations and modifications are available for students with disabilities, English language learners, and other students who need them in order to demonstrate their level of proficiency in the content area.
- Items and tasks are appropriate for the age and grade level of the students assessed.
- Items and tasks and the assessment as a whole have been reviewed for potential bias (including stereotypical and offensive content) against groups of students based on race, ethnicity, culture, language, religion, disabilities, gender, or region, etc.
- The assessment is free of irrelevant factors that are likely to interfere with students' opportunity to demonstrate their knowledge and skills, such as assumptions about background experiences and extraneous prior knowledge.
- Statistical item and task analyses (including bias analyses) indicate that all students have the opportunity to demonstrate their knowledge and skills.

Reporting

Reports of assessment results to the public, teachers, parents, students, and school administrators should be closely tied to content and performance standards. Different states have different purposes for

reporting and using reports. Score reports should be evaluated based on the purposes and uses of each report.

Considerations in evaluating reporting include the following:

- Score reports clearly illustrate levels of student proficiency on content standards.
- Reports contain information that can be used to make valid inferences and decisions.
- Reports contain information about the degree of uncertainty associated with reported scores (e.g., standard error of measure).
- Reports provide information that can be applied for the intended purpose(s) of the assessment, at the intended levels of aggregation (e.g., school, district).
- Scores are reported disaggregated by important categories of students (e.g., economic status, English proficiency, racial/ethnic group, gender, disability status).
- Reports are easily and appropriately interpreted by the intended audiences.

REFERENCES

Bailey, A. & Ross, G. (1998, April) Untitled presentation at a joint meeting of the Council of Chief State School Officers and the Council of the Great City Schools, Washington, DC.

Carlson, Dale (1996, October). *Adequate yearly progress in Title I of the Improving America's School Act: Issues and strategies*. Washington, DC: Council of Chief State School Officers.

Costa, A.L. (1989, April). Re-assessing assessment. *Educational Leadership, 46* (3).

Hansche, L.N. (1998). *Meeting the requirements of Title I: Handbook for the development of performance standards*. Washington, DC: U.S. Department of Education.

Kentucky Department of Education (1993, January). *Kentucky Instructional Results Information System: 1991-92 technical report*. Frankfort, KY: Author.

Kirst, Michael (1998). *Admissions and placement policies: Strengthening the K-16 Bridge*. National Center for Postsecondary Improvement Technical Report 2-06. Stanford, CA: NCPI.

Linn, Robert L. (1998). *Assessments and accountability*. CSE Technical Report 490. Los Angeles, CA: CRESST.

Long, V.M. & Benson, C. (1998, September). Re. Alignment. *The Mathematics Teacher, 91* (6), 503-508.

Maryland State Department of Education (1996, January). *1995 Maryland School Performance Assessment Program: Technical report*. Baltimore, MD: Author.

Mills, C.N., & Jaeger, R.M. (1998). Creating descriptions of desired student achievement when setting performance standards. In L. N. Hansche, *Meeting the Requirements of Title I: Handbook for the Development of Performance Standards*. Washington, DC: U.S. Department of Education.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards of school mathematics*. Reston, VA: Author.

National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.

National Research Council (1999). *Testing, teaching, and learning: A guide for states and school districts*. Washington, DC: National Academy Press.

Pipho, C. (1997, May). Stateline: Standards, assessment, accountability -- the tangled triumvirate. *Kappan, 78* (9), 673-674.

Romberg, T.A. & Carpenter, T.P. (1986). Research on teaching and learning mathematics: Two disciplines of scientific inquiry. In M.C. Wittrock (Ed.), *Handbook of Research on Teaching*. New York: Macmillan.

Romberg, T.A. & Wilson, L.D. (1995). Issues related to the development of an authentic assessment system for school mathematics. In T.A. Romberg (Ed.), *Reform in School Mathematics and Authentic Assessment*. Albany, N.Y.: State University of New York Press.

Sanford, E.E. & Fabrizio, L.M. (1999, April). *Results from the North Carolina-NAEP comparison and what they mean to the End-of-Grade Testing Program.* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec.

Schmidt, William (1999, June). Presentation in R. Blank (Moderator), *The alignment of standards and assessments*. Annual National Conference on Large-Scale Assessment, Snowbird, UT.

Stein, M.K., Grover, B.W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455-488.

Texas Education Agency, Student Assessment Division (1999, April 13). *1997-98 Technical digest.* Austin, TX: Author.

US Department of Education (1999, June, draft). *Peer reviewer guidance for evaluating evidence of final assessments under Title I.* Washington DC: Author.

Virginia Department of Education. (1993, June). *Virginia's guidelines for the review and development of assessment.* Richmond, VA: Author.

Walkup, Hugh (1999, January, draft). *Aligning assessment with learning: Testing for success.* Washington, DC: US Department of Education.

Webb, Norman L. (April, 1997a). *Research Monograph No. 6: Criteria for alignment of expectations and assessments in mathematics and science education.* Washington, DC: Council of Chief State School Officers.

Webb, Norman L. (January 1997b). Determining alignment of expectations and assessments in mathematics and science education. *National Institute for Science Education, Vol. 1*, No. 2.

Webb, Norman L. (1999a). *Alignment of science and mathematics standards and assessments in four states*. Washington, DC: Council of Chief State School Officers.

Webb, Norman L. (1999b). Presentation in R. Blank (Moderator), *The alignment of standards and assessments*. Annual National Conference on Large-Scale Assessment, Snowbird, UT.

Wixson, Karen (1999). Presentation in R. Blank (Moderator), *The alignment of standards and assessments*. Annual National Conference on Large-Scale Assessment, Snowbird, UT.

TM034236

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

NOTICE

Reproduction Basis

☒ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").